THERAPEUTIC ADVANCES in
Psychopharmacology

Review

History repeating: guidelines to address common problems in psychedelic science

Ther Adv Psychopharmacol

2023, Vol. 14: 1–20

DOI: 10.1177/20451253231198466

Michiel van Elk* and **Eiko I. Fried***

***Abstract*:** Research in the last decade has expressed considerable optimism about the clinical potential of psychedelics for the treatment of mental disorders. This optimism is reflected in an increase in research papers, investments by pharmaceutical companies, patents, media coverage, as well as political and legislative changes. However, psychedelic science is facing serious challenges that threaten the validity of core findings and raise doubt regarding clinical efficacy and safety. In this paper, we introduce the 10 most pressing challenges, grouped into easy, moderate, and hard problems. We show how these problems threaten internal validity (treatment effects are due to factors unrelated to the treatment), external validity (lack of generalizability), construct validity (unclear working mechanism), or statistical conclusion validity (conclusions do not follow from the data and methods). These problems tend to co-occur in psychedelic studies, limiting conclusions that can be drawn about the safety and efficacy of psychedelic therapy. We provide a roadmap for tackling these challenges and share a checklist that researchers, journalists, funders, policymakers, and other stakeholders can use to assess the quality of psychedelic science. Addressing today's problems is necessary to find out whether the optimism regarding the therapeutic potential of psychedelics has been warranted and to avoid history repeating itself.

Keywords: open science, psychedelics, psychotherapy, questionable research practices, validity

Received: 2 May 2023; revised manuscript accepted: 11 August 2023.

Correspondence to:
Michiel van Elk
Cognitive Psychology Unit, Institute of Psychology, Leiden University, PO Box 9555, Leiden 2300 RB, The Netherlands
m.van.elk@fsw.leidenuniv.nl

Eiko I. Fried
Clinical Psychology Unit, Institute of Psychology, Leiden University, Leiden, The Netherlands

*These authors contributed equally to this work.

Introduction

Psychedelics are mind-altering substances and include serotonergic hallucinogens [e.g. psilocybin, lysergic acid diethylamide (LSD), and Dimethyltryptamine (DMT)], entactogens [e.g. 3,4-methylenedioxy-methamphetamine (MDMA)], and dissociatives (e.g. ketamine). In the last decade, we have witnessed increased enthusiasm regarding the clinical application of psychedelics. Preliminary results from clinical trials using psychedelic therapy appear to show potential for the treatment of a wide variety of mental disorders, including major depressive disorder (MDD),[1] end-of-life-anxiety,[2] and addiction.[3] Results from phase II and phase III studies indicate that MDMA-assisted psychotherapy may be efficacious in the treatment of post-traumatic stress-disorder (PTSD).[4] And ketamine is increasingly being used in patients with treatment-resistant depression or suicidal ideation.[5,6]

Research also appears to show that psychedelics have beneficial effects in healthy volunteers. Participants in experimental research have rated psychedelic trips among their most meaningful life-events ever,[7,8] and psychedelic experiences in turn appear to cause other positive outcomes such as a healthier lifestyle,[9] increased mindfulness,[10] enhanced creativity and problem-solving,[11] pro-environmental behavior,[12] and feelings of connectedness.[13]

Taken together, the apparent benefits of psychedelics have led to significant changes in the mental health landscape. In the United States, more than 500 specialized 'ketamine-clinics' have

Correction (December 2023): This article has been updated with minor editing corrections in the 'Introduction' section and heading of first column in Table 1. Also, reference 'Marschall et al. 2022' has been added in the article, for further details please see <https://doi.org/10.1177/20451253231223609>.

emerged[14] that offer ketamine infusions to patients.[15] MDMA-assisted psychotherapy is awaiting approval by the FDA as medicine for the treatment of PTSD,[4] and the nonmedical use of psilocybin has already been legalized in Oregon[16]; other states are to follow suit.[17] Centers in Portugal, the Netherlands, and South America organize legal psilocybin or ayahuasca retreats for the treatment of MDD[18] and Australia officially recognized psychedelics as medicines in early 2023.[19] Pharmaceutical companies have made considerable investments into clinical research and filed patents on the production of psychedelics (e.g. for synthesizing a polymorph of psilocybin) as well as the therapeutic processes, for example, the training of therapists; cf. Smith and Appelbaum.[20] This has resulted in a flurry of different clinical applications, which can be broadly grouped as 'psychedelic therapy', of which psychedelic-assisted psychotherapy is the specific application of psychedelics embedded within a psychotherapeutic context.

Threats to validity in psychedelic research

However, we see many reasons to be seriously concerned about this optimism. In this paper, we highlight 10 key challenges that psychedelic science is currently faced with. These challenges threaten four types of validity[21,22] and cast serious doubt on the inferences that have been drawn in research carried out within the last decade. Threats to *internal validity* are best thought of as rival explanations, and raise doubts about whether a particular psychedelic intervention, rather than other factors, explains the results of a study. *External validity* refers to the generalizability of the research findings from the studied sample to the population of interest. *Construct validity* concerns the question how constructs in a study are operationalized, and what the exact working mechanisms are. Finally, *statistical conclusion validity* concerns the extent to which the conclusions, based on the data and statistical analyses, are warranted.

In psychedelic studies, internal validity is commonly threatened by the lack of control groups, the breaking blind problem, and placebo effects. The main sources of threats to external validity are low-powered studies and a strong selection bias in the inclusion of participants. Threats to construct validity include measurement problems as well as the lack of long-term treatment effects and mechanisms of action. Statistical conclusion validity is threatened by the multiple comparisons problem, conflicts of interest (COIs), outcome switching, and other questionable research practices. Of note, these validity threats are well-understood,[21] and it is therefore surprising that psychedelic science appears to ignore many of the lessons learned from decades of research. Much of the recent work is history repeating itself.

Previous reviews on psychedelic therapy have focused on the lack of sufficient safety data,[23] the lack of open science practices,[24] the breaking blind problem,[25] and the role of expectations and placebo effects.[26] In this paper, we take a broader perspective. First, we highlight the 10 most pertinent challenges and discuss how these problems threaten valid inferences. We conclude that because these validity threats often co-occur in psychedelic studies, it is premature to draw firm conclusions about the safety and efficacy of psychedelic therapy. Second, because solutions for some of these problems are more readily available than for others, we classify the problems as easy, moderate, or hard, depending on the methods and resources required to solve them. Third, we suggest how the problems can be mitigated, and consider solutions in terms of both scientific methods (e.g. more rigorous measures, statistics, and reporting) and resources (e.g. more funding, more therapists, more time, more participants). Finally, we summarize concrete recommendations for researchers and clinicians to move forward (see Table 1) and provide a resource list for reviewers, journalists, funders, and policymakers to facilitate the accurate assessment of the quality and rigor of psychedelic studies (see Table 2).

The easy problems

Problems that are relatively easy to address within psychedelic science include (1) invalid statistical inferences and questionable research practices, (2) COIs, and (3) safety and adverse events.

Invalid statistical inferences and questionable research practices

A conclusion is valid if the inferences follow from the evidence presented. There are two tiers of challenges here. First, authors in the psychedelic literature have regularly drawn conclusions that are either misleading or clearly contrast with the data. Four examples follow to showcase that this problem is common practice. Abbar *et al.*[27] found in a randomized controlled trial (RCT) comparing

Table 1. Easy, moderate, and hard problems in psychedelic science – and how to address them.

Problems	Solutions
Easy problems	
Invalid statistical inferences	Involve independent statistical experts in peer-review process More transparency through open science practices, including: Publish (anonymous) reviews alongside scientific papers Disclose all measures and statistical analyses Transparent preregistration of all studies Publish null results Investigate robustness of results (e.g. *via* multiverse analyses or many-analyst approaches) Share de-identified data
Conflicts of interest	Transparent reporting of all COIs in publicly accessible ways (i.e. not paywalled) Systematic instead of narrative reviews
Safety and adverse events	Report adverse events transparently and systematically Independent arbiters assess whether adverse events are related to the treatment Include safety/adverse events as primary or secondary outcomes
Moderate problems	
Lack of control groups	Fund and carry out studies with control groups; team science/multi-center collaborations can mitigate costs
Sample sizes	Fund and carry out properly powered studies; team science/multi-center collaborations can mitigate costs
Selection bias	Transparently disclose inclusion and exclusion criteria and full recruitment procedure Assess and report patient background characteristics that may influence treatment efficacy Use more representative samples
Study duration	Include long-term outcomes (⩾12 months) Use larger samples to mitigate attrition
Hard problems	
Breaking blind problem	Assess and report blinding efficacy Include an active control condition Recruit participants without prior psychedelic experiences Control for the breaking blind problem statistically

(Continued)

Table 1. (Continued)

Problems	Solutions
Placebo effects	Include a third study arm without any intervention to test against the placebo effect Measure expectations of patients and therapists Convey more realistic expectations (stop the hype)
Mechanisms of action	Carry out independent and high-powered replication studies Assess and report therapeutic alliance, efficacy of psychoplastogens, and efficacy of other techniques that induce an altered state

COIs, conflicts of interest.

Table 2. Checklist for assessing the quality and scientific rigor of psychedelic studies for mental health problems.

Criterion	Description
1. Valid Inferences	Is there sufficient transparency around data collection and statistical analyses, and are the conclusions supported by the evidence? Is there evidence that the treatment and not other factors (e.g. breaking the blind) explain the difference between the intervention and control group? Have independent reviewers with the relevant statistical expertise been assigned to review the manuscript? Are the reviews publicly available?
2. Conflicts of interest	Are potential COIs reported transparently in the paper? What is the nature of these COIs, and, in the presence of severe COIs, are there sufficient safeguards in place so that the findings can be considered trustworthy (e.g. preregistration of primary outcomes and statistical analyses)? Are all included measures fully disclosed and reported?
3. Safety and adverse events	Is it easy to find all relevant information regarding adverse events in the study? Is there an independent arbiter to decide whether an adverse event is related to the treatment? Is the psychotherapy component of the study standardized and fully described? Were trained therapist used to carry out treatments?
4. Control group	Is a control group included to address common validity threats such as placebo effects, expectancy effects, and regression to the mean? If no control group is included, are interpretations sufficiently careful?
5. Sample size	Is a power or sensitivity analysis provided, and does it include a justification of the minimum effect size of interest? Is the study sufficiently powered to detect a difference between intervention and control group (rather than powered against no effect at all)?
6. Selection bias	Does the studied sample differ from the population of interest? Is a statement about constraints on generalizability included? Is demographic (e.g. gender, age, socioeconomic background) and clinical (e.g. severity, comorbidities) information provided?
7. Study duration	Do scientists follow the patients for a sufficient time frame to justify the conclusion that successful treatment took place, that is, that people have returned to a normal level of symptom load, wellbeing, and functioning?
8. Breaking blind problem	Have efforts been made to minimize the risk of unblinding (e.g. by using active placebos)? Was masking efficacy (i.e. if blinding succeeded) assessed and reported?
9. Placebo effects	Does the study design account for placebo effects, for example, by comparing the intervention arm against a control group? Does the study include measures to assess patient and therapist's expectations about treatment outcomes both at the beginning and during the treatment?
10. Mechanisms of action	Are inferences regarding potential mechanisms of action supported by evidence? Are the data and materials available in a repository for replication and secondary analyses?

COIs, conflicts of interest.

ketamine against placebo that there was no persistent benefit of ketamine over placebo at the exit timepoint of the trial in week 6, but concluded in the abstract that 'ketamine [. . .] has persistent benefits for acute care in suicidal patients'. Ionescu *et al.*[28] found in an open-label ketamine study that only 2 of 14 patients show sustained improvement at 3-month follow-up (which may well be due to the placebo effect or other factors), but the title of the paper reads 'Rapid and *Sustained* Reductions in Current Suicidal Ideation' (our highlight). Palhano-Fontes *et al.*[29] concluded in their ayahuasca study ($n = 14$ treatment, $n = 15$ placebo) that 'blindness was adequately preserved', when *all* participants in the treatment group said they believed they had received ayahuasca, but *less than half* of participants in the placebo group said so. And Daws *et al.*[30] compared two treatment arms, including one using psilocybin-assisted psychotherapy, against each other, concluding that one treatment outperformed the other despite the lack of a statistically significant interaction term between the treatments. Journals, reviewers, funders, and scientific institutions need to hold authors accountable for such inferences.

The second tier of challenges is when questionable research practices – that usually go hand in hand with a lack of transparency – raise doubts about the validity of conclusions.[31–33] One common practice concerns flexibility regarding the analysis of primary outcome measures, which is common in psychedelic science. A recent phase II trial[34] administering psilocybin for treatment-resistant depression funded by a pharmaceutical company registered their primary outcome on clinicaltrials.gov for the time frame 'up to 12 weeks', which allows for degrees of freedom in obtaining statistically significant results and severely threatens valid inferences. Another study, using ketamine to treat depression[35] switched a secondary to a primary outcome a year after data collection started, and the publication contains no results at the time point of 2 weeks that can be found in the clinicaltrials.gov registration (https://osf.io/uhdrp). Yet another psilocybin trial for depression[30] analyzed a different outcome than registered on clinicaltrials.gov.[36] In all these cases, doubts remain whether equally positive results had occurred if stricter procedures would have been in place.

Another concern is related to multiple testing. Psychedelic studies often contain a flurry of different outcomes, including physiological, neural, cognitive, and self-report measures. This dramatically increases the chances of false positive findings when not dealt with appropriately. For instance, a recent paper on the efficacy of psychedelic therapy concluded that several secondary outcomes clearly favored psilocybin over escitalopram, but the lack of correction for multiple testing raises doubts about the validity of these findings.[1] In our own research on psilocybin microdosing, we included six different tasks, with multiple subcomponents and measures per task,[37] totaling the number of dependent measures to more than 20. We found effects of psilocybin microdosing on two outcomes. But because we had all outcomes preregistered transparently, and these two findings did not survive correcting for multiple testing, they likely reflected chance findings.

There are further problems, such as selectively removing outliers that result in significant findings, interpreting nonsignificant *p*-values in small samples as 'trends toward significance' (but not interpreting barely significant *p*-values as trends toward nonsignificance), and using one-sided tests over two-sided tests to obtain desired results,[30] discussed in Love.[38] All these problems jointly threaten both internal validity and statistical conclusion validity, as it remains unclear whether the conclusions are supported by the data and if other explanations might be more plausible for the effects observed. These concerns go hand in hand with evidence for substantial publication bias in the clinical trial literature for both pharmacological[39] and psychotherapeutic interventions.[40] This means that significant findings in trials are much more likely to be published in scientific journals, whereas nonsignificant findings are often not published, inflating meta-analytic effect sizes and threatening statistical conclusion validity.

We see two main ways forward. First, journals need to evaluate research papers more rigorously. This includes vetting by statistical and clinical trial experts without conflicts of interest. For the Daws *et al.*[30] paper referenced above, we applaud one of the reviewers for openly acknowledging that they did not catch a statistical problem, which may have impacted their review of the paper, and thereby the main conclusions of the paper.[38] To foster transparency and accountability, reviews should be published alongside papers, giving the research community insights as to the rigor of the review process. Journals must also compare the manuscript with the clinical trial protocol and statistical

analysis plan. Papers in which conclusions stand in contrast to the presented evidence should not be considered for publication. Second, researcher degrees of freedom and lack of transparency that foster questionable research practices can be reduced by adherence to best practices. These include publishing null-findings, for instance in the format of a *null results in brief* section[41]; reporting on all measures and tasks, preferably in the main paper rather than somewhere in Supplemental Materials; and *pre*registering research studies (including outcome measures and statistical analyses) to distinguish between exploratory and confirmatory findings.[42] The *registered reports* publication format offers a useful tool to increase transparency, by providing authors with peer-review feedback prior to conducting a study and (if accepted) a guaranteed publication even in case of null results.[43] Of note, there are already well-established best practices that have shown to increase transparency in the reporting of clinical trials,[44] such as the Consolidated Standards of Reporting Trials (CONSORT) reporting guidelines specifying 25 criteria.[45] Unfortunately, guidelines are not always taken as seriously as they should – for example, an audit of the clinical trial literature in clinical psychology indicated that (pre-)registrations were often incomplete and lacked sufficient information to be reproducible.[46] Overall, we deem solutions to these problems readily available; they primarily require implementing more rigorous scientific and statistical methods and transparent evaluating and reporting.

COIs with sponsored studies

Early research on psychedelics was often initiated by academic centers and supported by philanthropic organizations, for example, the Multidisciplinary Association For Psychedelic Studies (MAPS); cf. Mitchell *et al.*[4] More recently, pharmaceutical companies have become involved,[47] leading to financial COIs for some funders and authors. COIs by themselves are not a red flag, but conclusions from papers with COIs should be interpreted with extra care, given that they tend to go together with problematic practices. For instance, an analysis of 397 clinical trials in psychiatry has shown that studies reporting a COI were five times more likely to report positive results.[48] COIs are also pronounced in the psychedelic literature: a recent opinion paper on ketamine for treatment-resistant depression featured a COI section nearly five times as long as the paper's introduction section[49]: of 25 authors, 19 declared COIs, including patents for treating depression with ketamine. We are not convinced that the research community always takes best COI practices sufficiently seriously. For example, while the organizers for the psychedelic science 2023 conference (with over 300 speakers) originally followed recommended disclosure protocols,[50] by asking speakers to declare COIs in a dedicated presentation slide, slides were removed by the organizers before the talks and moved to the conference mobile app.

Moving forward, policy decisions should rely on systematic rather than narrative reviews, given that narrative reviews allow for more degrees of freedom to obtain desired results.[51] Industry-independent experts should be involved in all stages of the clinical trial process, including study design, data collection, analysis, writing, and peer review. Furthermore, we encourage all journals to publish COIs in a publicly accessible way, rather than putting them behind paywalls. COIs should be clearly communicated in presentations at scientific conferences as well, instead of being skimmed over or hidden in online apps. Transparency and open science tools provide scientific solutions and help safeguard against COI. These tools include: preregistration; blinding of group membership when conducting analyses; multiverse analyses, a method for assessing the robustness and boundary conditions of a specific effect[52]; and many-analyst approaches see for instance.[53] These solutions are all readily available and implementing them will increase statistical conclusion validity of psychedelic science. Making more public funding available for conducting independent research is another prerequisite to safeguard research against COIs.

Safety and adverse events

Sufficient information on potential risks, dangers, and adverse events is missing to draw conclusions that psychedelics are safe to use in the context of mental health treatments. For example, although esketamine (in the form of a nasal spray) was approved by the FDA for treatment-resistant depression in 2019, there is evidence that the approval process overlooked red flags,[14] and a meta-analysis from 2021 concluded that there was insufficient data regarding the long-term safety of ketamine.[54] A recent analysis found a systematic underreporting of serious adverse events studies on the safety and efficacy of esketamine in depression[55]: 41.5% of serious adverse

advents that were found on clinicaltrials.gov were not reported in published articles, and nearly all of them (88 out of 94) occurred in esketamine treatment arms, not placebo arms.

We see similar issues for other drugs now. Adverse events for therapy with MDMA and classical serotonergic hallucinogens are not systematically assessed and reported.[56] A straightforward definition of an adverse event in psychedelic research is missing, and standardized measurements and transparent reporting of adverse events are lacking. Most studies rely only on spontaneous reporting by patients or therapists, which in turn require a careful process of interpretation to assess whether the adverse event can be attributed to the psychedelic therapy specifically.

And adverse events do occur.[57] A recent clinical trial using psychedelic therapy[34] reported increased suicidal ideation and intentional self-injury in the 10 and 25 mg psilocybin group, whereas the control group who received 1 mg of psilocybin remained unaffected. In the same study, suicidal behaviors were observed in three patients, but authors concluded in a media report that 'these cases were probably random events and unrelated to the dose of psilocybin, which would have been fully cleared from the patients' bodies'.[58] This is not the most careful interpretation of the data, and we note that the opposite rationale is applied to successful treatments: people get better despite the drug being fully cleared from the patients' bodies. Sometimes, conclusions are radically opposed to presented evidence. In a placebo-controlled study using ayahuasca as an intervention for depression, 4 of 15 participants in the experimental group (i.e. ~27%) had to be hospitalized for 1 week 'due to presenting a more delicate condition'.[29] Nonetheless, authors concluded that the 'study brings new evidence supporting the *safety* and therapeutic value of psychedelics' (our highlight).

Psychedelic experiences can have short- and long-term adverse consequences. Short-term risks include the destabilizing effects that psychedelics can have through the experiences they can trigger, which can be difficult to handle both for the person and their therapist[59] – experiences that can result to everything from increased acute agitation to prolonged emotional dysregulation.[60] Many people also report adverse after-effects including recurrent hallucinations, increased anxiety, and physiological discomfort.[61] In the long term, such experiences can also cause ontological shocks, resulting in a dramatic shift in one's religious and spiritual worldviews.[62] In a recent article, a patient who had participated in a psilocybin study described their state of confusion, anxiety, and distress,[60] resulting in a long and desperate search including the use of spiritual practices, meditation techniques, and theology. This illustrates the dramatic effects that psychedelic therapy may have on some patients, and the need for careful spiritual, existential, religious, and theological support,[63] long after psychedelic sessions.[64]

Unfortunately, such support is typically not included in the current clinical trials which – often funded through pharmacological companies – strive for cost- and time-effective interventions. This is perhaps best summarized in the words of a patient receiving MDMA-assisted psychotherapy: '*(. . .) they tore open my chest, and they repaired the little damage in the heart there but then everyone just walked away from the table and my chest was still wide open*'.

Another obvious risk of psychedelic therapy concerns safety in the therapy room. Some proponents of psychedelic therapy have argued that clients regress during trips and require close physical content,[65] and many protocols include 'nurturing touch' as an important component during therapy. The MAPS study protocol, for example, writes that 'mindful use of touch can be an important catalyst to healing'.[66] But already since decades there is 'a lack of consensus about the use of touch and the complex ethical and clinical issues surrounding its use',[67] a crucially important topic, especially in psychedelic therapy given the highly suggestible and vulnerable state clients are in. Overall, manuals provide little guidance, leaving what is acceptable to the interpretation of the therapist.[57] A patient who took part in the phase II trial for MDMA-assisted psychotherapy for the treatment of PTSD (MAPS) reported several inappropriate encounters, including sexual abuse by the therapist,[68] casting doubt whether a safe and supportive therapeutic environment can currently be guaranteed.[57]

Overall, problems related to safety and adverse events are considerable, and without systematic and transparent reporting, it remains unclear for which patients psychedelic therapy provides a safe alternative and for whom it imposes a potential risk. More systematic and large-scale studies

about the prevalence and persistence of adverse effects are needed before psychedelics can be considered safe and efficacious treatments for mental health problems. Without such data, we have history repeating all over again, for example, with respect to the adverse effects of mindfulness-based interventions,[69] opioids,[70] or ketamine.[56] Given the profound consequences that psychedelics can have, researchers, funders, and regulators should consider including safety and adverse events as core outcome measures in clinical studies, rather than reporting them (if at all) in Supplemental Materials. Furthermore, instead of having the researchers decide whether an adverse event is related to a drug or not, such decisions should be made by independent arbiters. Finally, the psychotherapy and its components (e.g. acceptance and commitment approach; use of 'nurturing touch') applied within psychedelic therapy need to be standardized (see also section 'Mechanism remains unclear') and rigorously evaluated using the same stringent criteria for assessing safety and efficacy.[57]

These solutions are also relatively easy to implement and primarily require more rigorous scientific reporting.

The moderate problems

Whereas the easy problems should be relatively simple to address, the moderate problems – (1) the lack of (good) control groups, (2) sample sizes, (3) selection bias, and (4) lack of long-term follow measurements – require more effort. Solving the problem of sample size primarily requires more financial resources; for the other problems, scientific solutions will have to play a role as well.

Lack of control groups

Many clinical studies on the potential utility of psychedelics are open label, that is, without a control group. This is a typical design to start investigations into potential new treatments, but strongly limits what can be learned about the potential efficacy of interventions.[71] For instance, a recent open-label study followed 27 patients with MDD after two doses of psilocybin with supportive psychotherapy.[72] A year later, around 60% of the participants were no longer depressed. The authors concluded that there were 'substantial antidepressant effects of psilocybin-assisted therapy'. But without a control group, we don't know if the recovery was *due to* the intervention. To put the 60% into context, an analysis of 19 studies showed that the expected 12-month recovery rate of people with MDD who do not receive any treatment is around 53%.[73]

Demonstrating that mental disorder severity or diagnostic prevalence decline during treatments without comparing recovery rates to placebo does not establish evidence that treatments are efficacious.[25] This inference threat that is well-established for pharmacological and psychotherapeutic treatments of mental disorders is even more pronounced in psychedelic research, especially for the many naturalistic field studies conducted at psychedelic retreats.[74,75] In these studies, strong expectations, peer pressure, and the effect of a charismatic leader may further increase the placebo effect see also Plesa and Petranker.[76]

One related challenge that highlights the importance of a control group is regression to the mean, which 'decision-makers should always consider [. . .] to be a viable explanation for the observed change in an outcome in a pre-post study'.[77] Regression to the mean is a decrease in scores over time due to selecting people based on extreme values (such as the threshold on depression severity) when entering studies. This problem is common in clinical trials and psychedelic studies are no exception, where a particular score on mental health problems is often required for enrolment.

Taken together, these problems imply that an improvement of patients in psychedelic open-label studies cannot be causally attributed to the treatment itself. Including a control condition is a first, crucial step to increase internal validity. Several studies have implemented control groups, including active placebos, for example, diphenhydramine for the treatment of alcohol abuse disorder[3]; low-dose psychedelics,[2] or treatment as usual, for example, selective serotonin reuptake-inhibitors (SSRIs) for the treatment of major depression; cf. Carhart-Harris *et al.*[1] This design enables comparing two treatments over time (reflecting differences of treatments), rather than just the improvement of one group (confounded by, for example, placebo response and regression to the mean), although threats to internal validity still need to be managed, for example, the amount of contact a person has with a therapist should be the same in both treatment arms; cf. Kazdin.[21] The question about what constitutes a good

control condition in psychedelic therapy is also directly related to the hard problem of 'breaking blind', discussed in section 'The breaking blind problem'. Adding a control group will increase costs and time, which can be mitigated *via* team science, slow science, and multi-center studies, that is, different teams pooling resources to design and carry out more rigorous and better controlled studies. This solution can reduce many threats to internal validity and will make findings of psychedelic studies more robust.

Sample size

We would not finance or conduct a poll about who will win the next US presidency in 20 or 50 participants because such samples are not sufficiently informative regarding the general population we want to learn about. The same applies to clinical trials: researchers do not conclude that Alice and Bob responded well to ketamine, but that 'Ketamine is a safe and efficacious treatments for MDD',[78] requiring randomly drawn samples that sufficiently large to represent the target population of interest (in the case of the study of Bahji *et al.*, people diagnosed with depression). That is currently not the case, and problems of small samples (a threat to external validity) are further exacerbated by highly selective groups of patients (a threat for internal and construct validity).

The most recent meta-analysis of the seven available RCTs for psychedelic therapy for MDD and anxiety-related disorders showed that there was a significant reduction in symptoms up to 5 weeks following the intervention.[79] However, the number of participants per treatment arm was small, with more recent studies having somewhat higher power than early studies.[80] As underpowered studies typically result in imprecise and inflated effect sizes,[81] especially when questionable research practices and publication bias come into play, it is plausible that the initial efficacy of psychedelic therapy was overestimated. When larger preregistered studies will be conducted, it is likely that the resulting effect sizes will be lower than previously reported.

This is also relevant for understanding working mechanisms regarding the acute action of psychedelics, given that studies about neuroscientific theories are based on about a dozen neuroimaging datasets with very small samples (see also section 'Mechanism remains unclear'). Some of the most widely cited papers do not contain more than 10 participants, and a particular dataset by Carhart-Harris *et al.*[82] has been re-analyzed in numerous ways across 10 papers[83]; to the authors' credit: they were one of the few groups that actually openly shared their data thereby allowing other researchers to conduct secondary analyses. Threats to external validity are exacerbated by the large degrees of freedom in the analysis of neuroimaging data and the lack of a standardized analysis pipeline,[84] likely leading to many false positive results in the literature on neural mechanisms underlying psychedelics.

Small sample sizes primarily threaten external validity and statistical conclusion validity, as it remains uncertain whether findings generalize to the broader population. Although the problem of underpowered studies has been recognized for a long time, there are still strong incentives for conducting studies focused on discovery and explanation rather than study designs that critically evaluate the intended effects of a given therapy.[71] This can be mitigated somewhat if findings from small studies are reported on properly and if open science practices are implemented to prevent obtaining false positive results in underpowered samples.

Addressing these threats requires larger samples, which in turn will cost more resources. Larger samples will also help addressing the important question of what treatments work for whom. It could well be that psychedelic therapy selectively works for specific patient groups, such as patients primarily coping with depression triggered through a life-threatening illness,[85] and current samples are not sufficiently large to find such effects: in a recently pooled moderation analysis of 17 ketamine studies including more than 800 patients; however,[86] no consistent moderators were identified. This, in turn, highlights the need for more team science, multi-center and multi-lab studies to conduct few high-powered, preregistered studies, rather than many underpowered open-label studies that can fall prey to publication bias and other issues.

Selection bias

Participants learn about clinical trials from media reports, trial advertisement, and clinician referral (Muthukumaraswamy *et al.*[25]). This, along with prior experiences, introduces different types of selection biases. For example, inclusion criteria of

a recent study were right-handedness, mild-to-moderate depressive symptoms without psychotic features, abstinence from medication, drugs and alcohol (including tapering-off antidepressant medication, a challenge for many patients), good physical health, and no suicide risk.[87] Consequently, only very selective groups of patients are eligible for participating in these studies. The recent phase III study initiated by the MAPS included only 26% of 345 treatment-seeing participants,[4] and the rules by which they were picked across all the participating sites are not fully transparent.

What consequence can selection have? Compared to the general population of treatment-seeking individuals, participants in small psychedelic studies may be easier to treat, a common phenomenon for psychotherapeutic and pharmacological treatments as well. For example, when applying exclusion criteria of 161 antidepressant efficacy trials to 1271 in patients diagnosed with depression, between 76% and 99% of participants would have been removed, largely due to factors resulting in more severe psychopathology, for example, suicidal ideation, comorbidities.[88,89] This is important because cases with more severe or complicated psychopathology have worse treatment outcomes.[90] Participants enrolled in psychedelic studies are potentially also more motivated than the average treatment-seeking population, given that they often actively reach out to investigators such as in the MAPS study. Furthermore, participants who have prior positive experiences with psychedelics are more likely to participate in psychedelic trials. Prior experiences with psychedelics increase the likelihood of breaking blind cf. Aday *et al.*[26] and Carbonaro *et al.*,[91] which may in turn amplify observed treatment efficacy and reduce adverse events.

These and other selection biases lead to an overestimation of the psychedelic treatment effect, producing a substantial threat to internal and external validity. A related threat is the focus of psychedelic studies on White, Educated, Industrialized, Rich, Democratic (WEIRD) samples.[92–94] As a result, it remains unclear to what extent obtained findings generalize to the population at large. It could well be that psychedelics can help depressed patients to better deal with existential and meaning-related issues, but only if the basic needs for social security and safety are fulfilled. Furthermore, the clinical-medical approach to psychedelics may not fit well with the worldview and practices among some indigenous people and ethnic minorities; theories about the potential psychoplastogenic effects of psychedelics showcase a medical-mechanistic framing that can clash with traditional beliefs around psychedelics as plant medicines and a way to communicate with the divine.[95]

Despite common claims in the literature that psychedelic drugs work, for example, 'for MDD'[78] or 'for PTSD',[4] the combination of small samples and selection bias does not allow for such conclusions. Addressing selection biases requires better science: researchers need to measure and report more. In addition to clinical information such as diagnoses, severity, and comorbidities, other potential factors should be assessed and reported, including motivation and prior experience with psychedelics. Researchers also need to be more transparent about inclusion and exclusion criteria and should disclose the full study protocol, screening instruments, and the number of participants included and excluded. The problem of selection bias is also related to availability of resources; with more funding and personnel, it would be possible to recruit larger and more representative samples, including more difficult-to-reach populations.

Lack of long-term follow-up

Proponents of psychedelic therapy argue that one of the most important breakthroughs would be to provide efficacious treatments to people with severe and chronic mental health problems for whom gold-standard treatments have not been successful.[96,97] This explains why many studies have been conducted in treatment-resistant populations.[79] However, follow-up times in these studies cannot support claims of successful treatments. For example, a 2016 paper using ayahuasca as treatment for treatment-resistant depression[74] followed 29 participants for a total of 7 days, concluding that their research supports 'the safety and therapeutic value of psychedelics'. Participants had suffered from depression on average for 11 years, and a 7-day follow-up period is entirely insufficient to determine treatment efficacy. In the most recent meta-analysis on the effects of psychedelic therapy for depression and anxiety,[79] only three studies reported treatment outcomes up to 5–8 weeks following the intervention.

Studies using ketamine face similar challenges and have often measured outcomes 4 h, 24 h, and a few days after infusion.[98,99] It is therefore not surprising that a recent meta-analysis concluded

that 'long-term safety and efficacy [. . .] are yet to be investigated'.[100] We note that this reflects the state of the scientific literature *after* the FDA approved this drug for treating depression.

In summary, most studies in the psychedelic literature have at best demonstrated short-lived symptom relief, rather than successful treatment, contrary to claims popular in this literature. Given that over half of all cases of untreated depression remit spontaneously within 1 year[73] – depression is an episodic disease – it cannot be concluded without longer-term follow-ups that psychedelic therapy offers promising treatments for mental disorders. Australia now recognizes psychedelics as medicines, despite warnings from researchers in the field, including those involved in the largest trials: 'These treatments are not well established at all for a sufficient level of broad-scale implementation [. . .]. We've got no data on long-term outcomes at all, so that worries me a lot'.[19]

This challenge primarily threatens construct and internal validity, as it remains unclear whether psychedelic therapy directly targets mental health symptoms or whether there are rival explanations, for example, a positive afterglow following psychedelic experiences.[101] Therefore, studies need to implement more long-term follow-up measurements, preferably up to 1 year following the intervention, and larger samples are required to mitigate (selective) attrition. Such studies would be more expensive to conduct, but they are required to conclude with any confidence that psychedelic treatments add to the landscape of available treatments and are worth the associated risks.

The hard problems

The hard problems in psychedelic science include (1) the breaking blind problem, (2) placebo effects, and (3) the lack of a clearly specified mechanism. Solving these problems will require both rigorous scientific work and innovation, as well as more financial resources.

The breaking blind problem

RCTs are considered the gold standard in clinical psychology and psychiatry. They are usually double blind, meaning that neither participants nor researchers are aware of group membership. Unfortunately, blinding of participants and researchers is, depending on the particular psychedelic substance, either difficult or impossible. This is because the psychoactive effects of an active dose of most psychedelic substances become obvious to both the participant and the experimenter or clinician in about 30–60 min after intake. The *breaking blind problem* is therefore the rule rather than the exception in psychedelic trials.[25] For instance, in a recently published placebo-controlled study, 15 of 19 participants correctly guessed they had received a placebo, whereas 12 of 15 participants correctly guessed they receive an active dose after a medium- to high-dose psilocybin session;[102] unfortunately they did not report the clinicians' guesses about treatment allocation. A recent systematic review indicated that although most studies on psychedelic therapy were nominally blinded, participant blinding was only assessed in 8 out of 81 studies.[103]

This threatens internal validity and the conclusions that can be drawn based on clinical studies for several reasons. First, it is a threat to valid measurement on the side of the clinical interviewer because raters who know about group membership may no longer be unbiased when determining whether patients improved, for example, in their PTSD symptoms. This bias may be exacerbated when interviewers have personal positive experiences with psychedelics.[104] Second, breaking the blind adds a confound regarding attention and care: given that psychedelic psychotherapy is conceptualized as synergistic process that is catalyzed by the psychedelic substance, researchers and clinicians may implicitly provide more attention and put in more effort when they come to realize that the patient received an active dose compared to a placebo.[26] Third, it is a threat to valid measurement on the side of the patient, as it boosts expectancy and therefore placebo effects.

Despite these threats, the breaking blind problem is largely ignored in the literature, as evidenced by the fact that researchers rarely report on masking efficacy, that is, the extent to which blinding was preserved for systematic reviews, see Muthukumaraswamy *et al.*[25] and Nayak *et al.*[103] For example, the very recently published, first placebo-controlled, double-blind RCT for psilocybin-assisted therapy in MDD[87] went through considerable efforts to mask treatment allocation, but did not assess or report whether blinding was successful. However, given that the authors' analyses reveal massive differences between groups in their subjective experiences (including oceanic boundlessness, anxious ego dissolution, auditory alterations,

impaired cognition and control, and disembodiment), the study is clearly not double blind, and should not have been published as such. Even in the most recent phase III MDMA study for PTSD[4] blinding was not formally assessed, and the authors only report 'anecdotal' data, showing that participants guessed group membership correctly with over 90% accuracy.

Overall, we strongly recommend to formally assess masking efficacy in modern clinical trials, given that the breaking blind problem is a severe threat to valid inferences, and it only takes very little effort to assess and report. The problem cannot be easily solved and other treatments in which experiences form an integral part of the therapy (e.g. exposure therapy or music therapy) face similar challenges.[105] Different solutions have been proposed to reduce the risk of breaking blind, including the use of active placebos that induce physiological, for example, such as methylphenidate; see for instance Griffiths *et al.*[8] or psychological effects, for example, comparing MDMA with psilocybin; cf. Muthukumaraswamy *et al.*[25] Recruiting psychedelic- naïve participants for clinical trials may reduce the risk of breaking blind somewhat for low-dose trials, as people without prior experience will find it more difficult to correctly guess their condition assignment.[26,91] Using different dosing conditions of the same substance (e.g. comparing 10, 20, and 30 mg of psilocybin) appears an effective way to avoid the breaking blind problem.[25,106] This design rests on the assumption that a low dose is not or less therapeutically effective, and it allows directly assessing the dose–response relationship. Selective or partial disclosure of information prior to the study could be another way to reduce the risk of breaking blind, although it faces ethical challenges.[26] For instance, providing instructions that the study consists of multiple arms (instead of two) with different dosing regimens reduces chances of breaking blind. Patients could also be informed about all possible side effects of all potential study drugs included in the clinical trial (i.e. rather than the side effects of each specific substance they are about to receive), to further minimize confidence about treatment allocation.[107] Alternatively, using a balanced placebo-informed design, drug information, and actual drug administration could be independently manipulated to allow a direct comparison of the effect of expectations about condition assignment and the actual psychedelic substance.[25]

A radical but ethically challenging idea was proposed by Nautiyal and Yaden[108]: in order to completely rule out the breaking-blind problem, patients could be anesthetized prior to the administration of psilocybin or placebo. In line with this suggestion, a recently registered study protocol intends to combine psilocybin with the benzodiazepine midazolam to induce amnesia for the psychedelic experience.[109] Similarly, in a recent study using a triple-masked randomized placebo-controlled design, MDD patients received a ketamine infusion or a placebo while being anesthetized, thereby effectively avoiding participants from breaking blind.[110] However, both the ketamine and the placebo group showed an antidepressant effect, which might be related to a general placebo response in both groups and/or the antidepressant effects of being anesthetized .

At the minimum, all clinical trials should always assess and disclose the rate of breaking blind: transparency is a first crucial step to get an idea of how severe the problem is. Another easy-to-implement solution is to use blinded analysis of the primary outcome measures,[111] or to keep raters blind to the full design of the study including which compound is being tested for further recommendations, see Even *et al.*[112] We see some progress in this area, given that new statistical tools are under development aiming to 'control for' participants who broke the blind during the study, and a new scale has been developed to formally assess whether participants broke blind.[113]

Placebo effects

Prior expectations about the efficacy of treatments have strong effects on participant experience and outcome reporting.[26] In the realm of psychedelic drugs, both the set (i.e. the person's expectations, bodily, and mental state) and the setting (i.e. the physical context and environmental factors) strongly contribute to the psychedelic experience, amplifying the placebo response.[114,115] This effect may be further enhanced because psychedelics increase suggestibility[116] and thus potentially function as a super placebo.[114] A recent meta-analysis indicated that efficacy of cannabis-based treatments for pain does not exceed placebo effects, and that media-reporting about the efficacy of cannabis is often inaccurate and overly positive, leading to exaggerated expectations by patients.[117] Similarly, given the current hype

around psychedelics, the so-called Pollan-effect reflects the exaggerated expectations that have been caused by selective media-reporting about the beneficial effects of psychedelics.[26] We see several challenges posed by this hype.

First, when participants in the placebo group know about the group membership (e.g. because they do not experience the psychedelic experiences they expect), this can lead to demotivation, nocebo effects, and dropout.[118] In the face of disappointment, an intervention group may outperform a placebo group not because the treatment works better than placebo, but because the control group works worse than placebo. Second, broken expectations can lead to false statements that are under-researched. Interviews with participants of the MAPS study have indicated that even though some patients felt worse following the treatment, they ended up reporting improvements. Participants felt immense pressure to report positive outcomes because investigators and media outlets had touted MDMA-assisted psychotherapy as highly efficacious treatments. Some participants said they were worried that honestly reporting their experiences might jeopardize approval of treatments that they understood to be potentially lifesaving for many others.[119]

Placebo effects threaten internal validity of treatments, as it remains unclear whether symptom improvement relies on the placebo effect or can truly be attributed to the effects of the psychedelic therapy. Including a third study arm (next to an active placebo), consisting of an additional control condition that receives no treatment at all, allows a formal quantification of the size of the placebo effect and a control for treatment nonspecific factors.[120] All studies should assess patients' and therapists' expectations prior to and during trials, which provides information necessary to control for expectation and placebo effects.[26] Another solution is to induce more realistic expectations for clinicians and patients by providing explicit information about the current uncertainty regarding the efficacy of psychedelic therapy in the informed consent of studies. More generally, we hope to see more nuanced conclusions in scientific studies, university press releases, and media reports; the current hype directly contributes to validity threats for psychedelic science and does a strong disservice to treatment-seeking populations.

Mechanism remains unclear

Several different working mechanisms have been proposed to account for the effects of psychedelics, including increases in neuroplasticity, neural entropy, or psychological flexibility.[121] Although some have argued that the subjective experiences are the primary mechanism of action,[122] others have suggested the neurochemical effects instead explains efficacy, such as potential neuroplastic effects of psychedelics, that is, the so-called neuroplastogen model.[123] A third model posits that psychedelics loosen maladaptive prior beliefs and increase sensitivity to bottom-up prediction error signaling.[124] On this account, the hallmark of most psychopathological disorders is thought to be rigidity in thinking, emotions and behavior, and psychedelics supposedly counter this rigidity, reflected in increased brain entropy following psychedelic therapy.[30]

But psychedelic therapy also consists of a psychotherapy component. This component is under-researched, with only 3 out of 21 studies having evaluated the effect of providing psychological support compared to a minimally supportive condition.[103] A related practical challenge for evaluating the efficacy of treatments is the wide variety of different therapeutic approaches used interchangeably in psychedelic therapy, often in the same study. This includes emotion-focused, psychodynamic, transpersonal, existential, or nondirective therapies,[125] but the MAPS protocol also lists internal family systems, voice dialog, psychosynthesis, Hakomi, sensorimotor therapy, and holotropic breathwork.[66] This in turn relates to the lack of standardized training and requirements for therapists: there are currently no formal minimum requirements for professional psychiatrists or psychologists to apply psychedelic therapy.

All proposed mechanisms and therapies have in common that there is a high degree of uncertainty and a lack of strong empirical evidence.[121,126] A parallel may be drawn with research on antidepressants: the serotonin hypothesis remains one of the most widely accepted mechanisms underlying the efficacy of SSRIs, but a recent systematic review indicated that there is little empirical support for this hypothesis as core working mechanisms for SSRIs.[127] Much more rigorous and fundamental research is necessary to establish the causal pathways through which psychedelics exert their potential effects to avoid the situation that the SSRI literature finds itself in, where proposed

working mechanisms may have been held up in part by false positive findings, an over simplistic narrative convenient for some stakeholders, and potential conflicts of interests.

Together, the lack of a clearly specified mechanism poses a threat to construct validity. Independent and high-powered replication studies are needed to test central predictions from promising neurobiological models specifying the causal-mechanistic pathways of psychedelics,[121] and work on the interaction between psychedelic use and psychotherapeutic interventions is crucial to move the field forward. In line with the observed importance of therapeutic alliance for predicting clinical improvements,[128] including an assessment of the relationship between the client and the therapist will provide insight in the psychotherapeutic component as a crucial mediating factor in psychedelic-assisted psychotherapy. The development of new classes of psychedelics, so-called psychoplastogens,[126] which induce neurobiological plasticity without the accompanying psychedelic effects, will allow assessing the importance of the different pharmacological and neural pathways that might contribute to therapeutic efficacy. And clinical research using other methods, such as meditation, sensory deprivation, or breathwork exercises, will shed light on the therapeutic importance of subjective experiences in the therapeutic process.[108,122]

Connecting the dots

Validity threats are problematic when they occur in isolation, but valid inferences become exponentially more difficult when validity threats interact with each other. In most psychedelic trials, the lack of appropriate controls, the breaking-blind problem, as well as expectancy and placebo effects typically tend to co-occur.[26] These problems become even more challenging in small and selective samples relying on short-term follow-ups, and on top of that, conflicts of interest come into play, which interact with the lack of proper scientific practices such as transparent preregistration of all measures and analyses, and publishing all findings irrespective of whether they are positive or not. These challenges, together with the lack of safety data, such as reliable and transparent reporting of adverse events, lead to the conclusion that it is too early to draw any firm conclusions regarding the efficacy and safety of psychedelic therapy.

These problems are not unique to psychedelic science and tend to affect treatments in clinical psychology and psychiatry as a whole. But the challenges have been recognized for a long time see for example Sterling *et al.*,[129] and numerous solutions are readily available, which as of yet have not been implemented by the psychedelic research community. Psychedelic science is history repeating, and little concerted action has been taken to reduce bias in reporting and publishing of results.

Conclusion: A roadmap for psychedelic science

To improve the rigor and credibility of psychedelic clinical science, we need to set up studies aiming to address as many validity threats as possible. In Table 2, we provide a brief checklist that can be used by researchers, funders, reviewers, and policymakers to vet the quality of psychedelic studies see also Schiavone *et al.*[130] Such criteria can help determining the value of previously conducted work but can also shape the future. Researchers, funders, and policymakers can use this list to assess if planned studies are fit to meet the needs of psychedelic science moving forward. Note that these criteria need to be interpreted in the context of the study, of course; meeting all criteria does not necessarily make a study rigorous, and many studies not ticking all boxes will be valuable, too. NIH recently developed guidelines for funding psychedelic research, concluding that studies that lack 'basic quality controls and methodological rigor' should be considered as 'low priority'.[131] In our reading, this renders nearly all work in this field as 'low priority'.

Psychedelic science can move forward by including appropriate control conditions, more participants, more diverse samples, long-term follow-up measurements, and by being a lot more transparent about measures and methods. The hard problems require more rigorous scientific work, but even here, transparency is a necessary first step: the number of people breaking blind, expectations about the treatment, adverse events, and other information we discussed should be collected and reported.

Given the current state of research, strong caution is warranted regarding the hype around psychedelics as treatments: there is not enough robust evidence to draw any firm conclusions about the safety and efficacy of psychedelic therapy. Our hope is that new studies may find credible evidence that psychedelic therapy can be a useful tool for the treatment of specific groups of patients. Until that time, we urge caution repeating the history of so many hyped treatments in clinical psychology and psychiatry in the last century. For psychedelic research in particular, we are not the first to raise concerns and can only echo the warning expressed more than half a century ago:

> To be hopeful and optimistic about psychedelic drugs and their potential is one thing; to be messianic is another. Both the present and the future of psychedelic research already have been grievously injured by a messianism that is as unwarranted as it has proved undesirable.[132]

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Author contributions
Michiel van Elk: Conceptualization; Formal analysis; Funding acquisition; Project administration; Writing – original draft; Writing – review & editing.

Eiko I. Fried: Conceptualization; Formal analysis; Methodology; Writing – original draft; Writing – review & editing.

Acknowledgements
We would like to thank colleagues and people who provided their valuable feedback on earlier versions of this manuscript.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MvE was supported to conduct this research by a VIDI grant from the Netherlands Organization for Scientific Research (NWO; grant id# VI.Vidi.191.107). EF did not receive funding for this research. Our study did not require an ethical board approval because it did not contain animal or human trials.

Competing interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This project was supported by a VIDI grant from the Netherlands Organization for Scientific Research (NWO; grant id# VI.Vidi.191.107).

Availability of data and materials
Not applicable.

ORCID iD
Michiel van Elk https://orcid.org/0000-0002-7631-3551

References

1. Carhart-Harris R, Giribaldi B, Watts R, *et al.* Trial of psilocybin versus escitalopram for depression. *N Engl J Med* 2021; 384: 1402–1411.
2. Griffiths RR, Johnson MW, Carducci MA, *et al.* Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. *J Psychopharmacol* 2016; 30: 1181–1197.
3. Bogenschutz MP, Ross S, Bhatt S, *et al.* Percentage of heavy drinking days following psilocybin-assisted psychotherapy vs placebo in the treatment of adult patients with alcohol use disorder: a randomized clinical trial. *JAMA Psychiatry* 2022; 79: 953–962.
4. Mitchell JM, Bogenschutz M, Lilienstein A, *et al.* MDMA-assisted therapy for severe PTSD: a randomized, double-blind, placebo-controlled phase 3 study. *Nat Med* 2021; 27: 1025–1033.
5. Canuso CM, Singh JB, Fedgchin M, *et al.* Efficacy and safety of intranasal esketamine for the rapid reduction of symptoms of depression and suicidality in patients at imminent risk for suicide: results of a double-blind, randomized, placebo-controlled study. *Focus* 2019; 17: 55–65.
6. Veraart JKE, Smith-Apeldoorn SY, Spaans HP, *et al.* Is ketamine an appropriate alternative to ECT for patients with treatment resistant depression? A systematic review. *J Affect Disord* 2021; 281: 82–89.

7. Griffiths RR, Johnson MW, Richards WA, *et al.* Psilocybin occasioned mystical-type experiences: immediate and persisting dose-related effects. *Psychopharmacology* 2011; 218: 649–665.
8. Griffiths RR, Richards WA, McCann U, *et al.* Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology* 2006; 187: 268–283.
9. Teixeira P, Johnson M, Timmermann C, *et al.* Psychedelics and health behavior change. *J Psychopharmacol* 2022; 36: 12–19.
10. Smigielski L, Kometer M, Scheidegger M, *et al.* Characterization and prediction of acute and sustained response to psychedelic psilocybin in a mindfulness group retreat. *Sci Rep* 2019; 9: 14914.
11. Girn M, Mills C, Roseman L, *et al.* Updating the dynamic framework of thought: creativity and psychedelics. *Neuroimage* 2020; 213: 116726.
12. Forstmann M and Sagioglou C. Lifetime experience with (classic) psychedelics predicts pro-environmental behavior through an increase in nature relatedness. *J Psychopharmacol* 2017; 31: 975–988.
13. Carhart-Harris RL, Erritzoe D, Haijen E, *et al.* Psychedelics and connectedness. *Psychopharmacology* 2018; 235: 547–550.
14. Strakowski SM, Sanacora G and Nemeroff CB. Does ketamine live up to the hype in depression?, https://www.medscape.com/viewarticle/976364 (2023, accessed 18 September 2023).
15. Mathai DS, Lee SM, Mora V, *et al.* Mapping consent practices for outpatient psychiatric use of ketamine. *J Affect Disord* 2022; 312: 113–121.
16. Jacobs A. Legal use of hallucinogenic mushrooms begins in Oregon, https://www.nytimes.com/2023/01/03/health/psychedelic-drugs-mushrooms-oregon.html (2023, accessed 18 September 2023).
17. Marks M and Cohen IG. Patents on psychedelics: the next legal battlefront of drug development. *Harv Law Rev Forum* 2021; 135: 212.
18. Rucker JJ and Young AH. Psilocybin: from serendipity to credibility? *Front Psychiatry* 2021; 12: 659044.
19. Chrysanthos N and Dow A. Australia becomes first country to recognise psychedelics as medicines, https://www.smh.com.au/politics/federal/australia-becomes-first-country-to-recognise-psychedelics-as-medicines-20230203-p5chs6.html (2023, accessed 18 September 2023).
20. Smith WR and Appelbaum PS. Novel ethical and policy issues in psychiatric uses of psychedelic substances. *Neuropharmacology* 2022; 216: 109165.
21. Kazdin AE. *Research design in clinical psychology*. Cambridge, UK: Cambridge University Press, 2021.
22. Shadish WR, Cook TD and Campbell DT. Statistical conclusion validity and internal validity. In: Cook TD, Campbell DT and Shadish W (eds) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin, 2002, pp. 103–134.
23. Sanacora G, Frye MA, McDonald W, *et al.* A consensus statement on the use of ketamine in the treatment of mood disorders. *JAMA Psychiatry* 2017; 74: 399–405.
24. Petranker R, Anderson T and Farb N. Psychedelic research and the need for transparency: polishing Alice's looking glass. *Front Psychol* 2020; 11: 1681.
25. Muthukumaraswamy SD, Forsyth A and Lumley T. Blinding and expectancy confounds in psychedelic randomized controlled trials. *Expert Rev Clin Pharmacol* 2021; 14: 1133–1152.
26. Aday JS, Heifets BD, Pratscher SD, *et al.* Great Expectations: recommendations for improving the methodological rigor of psychedelic clinical trials. *Psychopharmacology* 2022; 239: 1989–2010.
27. Abbar M, Demattei C, El-Hage W, *et al.* Ketamine for the acute treatment of severe suicidal ideation: double blind, randomised placebo controlled trial. *BMJ* 2022; 376: e067194.
28. Ionescu DF, Swee MB, Pavone KJ, *et al.* Rapid and sustained reductions in current suicidal ideation following repeated doses of intravenous ketamine: secondary analysis of an open-label study. *J Clin Psychiatry* 2016; 77: e719–e725.
29. Palhano-Fontes F, Barreto D, Onias H, *et al.* Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychol Med* 2019; 49: 655–663.
30. Daws RE, Timmermann C, Giribaldi B, *et al.* Increased global integration in the brain after psilocybin therapy for depression. *Nat Med* 2022; 28: 844–851.
31. Banks GC, Rogelberg SG, Woznyj HM, *et al.* Evidence on questionable research practices: the good, the bad, and the ugly. *J Bus Psychol* 2016; 31: 323–338.
32. John LK, Loewenstein G and Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 2012; 23: 524–532.
33. NWO. Netherlands code of conduct for research integrity, https://www.nwo.nl/en/

netherlands-code-conduct-research-integrity (2018, accessed 18 September 2023).

34. Goodwin GM, Aaronson ST, Alvarez O, *et al*. Single-dose psilocybin for a treatment-resistant episode of major depression. *N Engl J Med* 2022; 387: 1637–1648.
35. Price RB, Spotts C, Panny B, *et al*. A novel, brief, fully automated intervention to extend the antidepressant effect of a single ketamine infusion: a randomized clinical trial. *Am J Psychiatry* 2022; 179: 959–968.
36. Doss MK, Barrett FS and Corlett PR. Skepticism about recent evidence that psilocybin "liberates" depressed minds. *ACS Chem Neurosci* 2022; 13: 2540–2543.
37. Marschall J, Fejer G, Lempe P, *et al*. Psilocybin microdosing does not affect emotion-related symptoms and processing: a preregistered field and lab-based study. *J Psychopharmacol* 2022; 36: 97–113.
38. Love S. Inside the dispute over a high-profile psychedelic study, https://www.vice.com/en/article/4awj3n/inside-the-dispute-over-a-high-profile-psychedelic-study (2022, accessed 18 September 2023).
39. Turner EH, Matthews AM, Linardatos E, *et al*. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; 358: 252–260.
40. Driessen E, Hollon SD, Bockting CL, *et al*. Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US National Institutes of Health-Funded Trials. *PLoS One* 2015; 10: e0137864.
41. Munafo M and Neill J. Null is beautiful: on the importance of publishing null results. *J Psychopharmacol* 2016; 30: 585.
42. Claesen A, Gomes S, Tuerlinckx F, *et al*. Comparing dream to reality: an assessment of adherence of the first generation of preregistered studies. *R Soc Open Sci* 2021; 8: 211037.
43. Chambers CD and Tzavella L. The past, present and future of Registered Reports. *Nat Hum Behav* 2022; 6: 29–42.
44. Hopewell S, Ravaud P, Baron G, *et al*. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012; 344: e4178.
45. Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.
46. Cybulski L, Mayo-Wilson E and Grant S. Improving transparency and reproducibility through registration: the status of intervention trials published in clinical psychology journals. *J Consult Clin Psychol* 2016; 84: 753–767.
47. Goodwin GM, Aaronson ST, Alvarez O, *et al*. Single-dose psilocybin for a treatment-resistant episode of major depression: impact on patient-reported depression severity, anxiety, function, and quality of life. *J Affect Disord* 2023; 327: 120–127.
48. Perlis RH, Perlis CS, Wu Y, *et al*. Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *Am J Psychiatry* 2005; 162: 1957–1960.
49. McIntyre RS, Rosenblat JD, Nemeroff CB, *et al*. Synthesizing the evidence for ketamine and esketamine in treatment-resistant depression: an international expert opinion on the available evidence and implementation. *Am J Psychiatry* 2021; 178: 383–399.
50. Martin SJ and Hunt DPJ. *Assessment of comprehensibility of industry conflicts of interest and disclosures by multiple sclerosis researchers at medical conferences*. JAMA Netw Open 2021; 4: e212167.
51. Thomas-Odenthal F, Molero P, van der Does W, *et al*. Impact of review method on the conclusions of clinical reviews: a systematic review on dietary interventions in depression as a case in point. *PLoS One* 2020; 15: e0238131.
52. Steegen S, Tuerlinckx F, Gelman A, *et al*. Increasing transparency through a multiverse analysis. *Perspect Psychol Sci* 2016; 11: 702–712.
53. Hoogeveen S, Sarafoglou A, Aczel B, *et al*. A many-analysts approach to the relation between religiosity and well-being. *Religion Brain Behav*. Epub ahead of print 6 July 2022. DOI: 10.1080/2153599X.2022.2070255.
54. Ceban F, Rosenblat JD, Kratiuk K, *et al*. Prevention and management of common adverse effects of ketamine and esketamine in patients with mood disorders. *CNS Drugs* 2021; 35: 925–934.
55. de Laportaliere T, Jullien A, Yrondi A, *et al*. Reporting of harms in clinical trials of esketamine in depression: a systematic review. *Psychol Med.* Epub ahead of print 26 April 2023. DOI: 10.1017/S0033291723001058.
56. Breeksema JJ, Kuin BW, Kamphuis J, *et al*. Adverse events in clinical treatments with serotonergic psychedelics and MDMA: a mixed-methods systematic review. *J Psychopharmacol* 2022; 36: 1100–1117.
57. McNamee S, Devenot N and Buisson M. Studying harms is key to improving psychedelic-assisted

therapy – participants call for changes to research landscape. *JAMA Psychiatry* 2023; 80: 411–412.

58. Sample I. Magic mushrooms' psilocybin can alleviate severe depression when used with therapy, https://www.theguardian.com/science/2022/nov/02/magic-mushrooms-psilocybin-alleviate-severe-depression-alongside-therapy (2022, accessed 18 September 2023).
59. Oehen P and Gasser P. Using a MDMA- and LSD-group therapy model in clinical practice in switzerland and highlighting the treatment of trauma-related disorders. *Front Psychiatry* 2022; 13: 863552.
60. Petersen R. A theological reckoning with 'bad trips', https://bulletin.hds.harvard.edu/a-theological-reckoning-with-bad-trips/#Notes (2023, accessed 18 September 2023).
61. Barrett FS, Bradstreet MP, Leoutsakos JS, *et al.* The Challenging Experience Questionnaire: characterization of challenging experiences with psilocybin mushrooms. *J Psychopharmacol* 2016; 30: 1279–1295.
62. Timmermann C, Kettner H, Letheby C, *et al.* Psychedelics alter metaphysical beliefs. *Sci Rep* 2021; 11: 22166.
63. Palitsky R, Kaplan DM, Peacock C, *et al.* Importance of integrating spiritual, existential, religious, and theological components in psychedelic-assisted therapies. *JAMA Psychiatry* 2023; 80: 743–749.
64. Bathje GJ, Majeski E and Kudowor M. Psychedelic integration: an analysis of the concept and its practice. *Front Psychol* 2022; 13: 824077.
65. McLane H, Hutchison C, Wikler D, *et al.* Respecting autonomy in altered states: navigating ethical quandaries in psychedelic therapy. *J Med Ethics* 2021.
66. Mithoefer MC. *A manual for MDMA-assisted psychotherapy in the treatment of posttraumatic stress disorder*. Multidisciplinary Association for Psychedelic Studies (MAPS): Santa Cruz, CA, 2015.
67. Durana C. The use of touch in psychotherapy: ethical and clinical guidelines. *Psychotherapy* 1998; 35: 269–280.
68. Lindsay B. Footage of therapists spooning and pinning down patient in B.C. trial for MDMA therapy prompts review, https://www.cbc.ca/news/canada/british-columbia/bc-mdma-therapy-videos-1.6400256 (2022, accessed 18 September 2023).
69. Britton WB, Lindahl JR, Cooper DJ, *et al.* Defining and measuring meditation-related adverse effects in mindfulness-based programs. *Clin Psychol Sci* 2021; 9: 1185–1204.
70. Naudet F, Fried EI, Cosgrove L, *et al.* Psychedelic drugs: more emphasis on safety issues. *Nature* 2022; 611: 449.
71. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med* 2008; 5: e67.
72. Gukasyan N, Davis AK, Barrett FS, *et al.* Efficacy and safety of psilocybin-assisted treatment for major depressive disorder: prospective 12-month follow-up. *J Psychopharmacol* 2022; 36: 151–158.
73. Whiteford HA, Harris MG, McKeon G, *et al.* Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychol Med* 2013; 43: 1569–1585.
74. Sanches RF, de Lima Osorio F, Dos Santos RG, *et al.* Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: a SPECT study. *J Clin Psychopharmacol* 2016; 36: 77–81.
75. van Oorsouw K, Toennes SW and Ramaekers JG. Therapeutic effect of an ayahuasca analogue in clinically depressed patients: a longitudinal observational study. *Psychopharmacology* 2022; 239: 1839–1852.
76. Plesa P and Petranker R. Manifest your desires: psychedelics and the self-help industry. *Int J Drug Policy* 2022; 105: 103704.
77. Linden A. Assessing regression to the mean effects in health care initiatives. *BMC Med Res Methodol* 2013; 13: 119.
78. Bahji A, Zarate CA and Vazquez GH. Efficacy and safety of racemic ketamine and esketamine for depression: a systematic review and meta-analysis. *Expert Opin Drug Saf* 2022; 21: 853–866.
79. Ko K, Kopra EI, Cleare AJ, *et al.* Psychedelic therapy for depressive symptoms: a systematic review and meta-analysis. *J Affect Disord* 2023; 322: 194–204.
80. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294: 218–228.
81. Johnsen TJ and Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. *Psychol Bull* 2015; 141: 747–768.
82. Carhart-Harris RL, Erritzoe D, Williams T, *et al.* Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proc Natl Acad Sci U S A* 2012; 109: 2138–2143.

83. McCulloch DE, Knudsen GM and Fisher PMP. Sychedelic resting-state neuroimaging: a review and perspective on balancing replication and novel analyses. *PsyArXiv*, 2021.

84. Vul E, Harris C, Winkielman P, *et al*. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci* 2009; 4: 274–290.

85. Yaden DB, Nayak SM, Gukasyan N, *et al*. The potential of psychedelics for end of life and palliative care. *Curr Top Behav Neurosci* 2022; 56: 169–184.

86. Price RB, Kissel N, Baumeister A, *et al*. International pooled patient-level meta-analysis of ketamine infusion for depression: in search of clinical moderators. *Mol Psychiatry* 2022; 27: 5096–5112.

87. von Rotz R, Schindowski EM, Jungwirth J, *et al*. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. *EClinicalMedicine* 2023; 56: 101809.

88. Zimmerman M, Balling C, Chelminski I, *et al*. Applying the inclusion/exclusion criteria in placebo-controlled studies to a clinical sample: a comparison of medications. *J Affect Disord* 2020; 260: 483–488.

89. Zimmerman M, Chelminski I and Posternak MA. Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. *Am J Psychiatry* 2005; 162: 1370–1372.

90. Kim HK, Blumberger DM, Fitzgerald PB, *et al*. Antidepressant treatment outcomes in patients with and without comorbid physical or psychiatric disorders: a systematic review and meta-analysis. *J Affect Disord* 2021; 295: 225–234.

91. Carbonaro TM, Johnson MW, Hurwitz E, *et al*. Double-blind comparison of the two hallucinogens psilocybin and dextromethorphan: similarities and differences in subjective experiences. *Psychopharmacology* 2018; 235: 521–534.

92. George DR, Hanson R, Wilkinson D, *et al*. Ancient roots of today's emerging renaissance in psychedelic medicine. *Cult Med Psychiatry* 2022; 46: 890–903.

93. Breeksema JJ and van Elk M. Working with weirdness: a response to "moving past mysticism in psychedelic science". *ACS Pharmacol Transl Sci* 2021; 4: 1471–1474.

94. Michaels TI, Purdon J, Collins A, *et al*. Inclusion of people of color in psychedelic-assisted psychotherapy: a review of the literature. *BMC Psychiatry* 2018; 18: 245.

95. Hartogsohn I. Set and setting in the Santo Daime. *Front Pharmacol* 2021; 12: 651037.

96. Carhart-Harris RL and Goodwin GM. The therapeutic potential of psychedelic drugs: past, present, and future. *Neuropsychopharmacology* 2017; 42: 2105–2113.

97. Curran HV, Nutt D and de Wit H. Psychedelics and related drugs: therapeutic possibilities, mechanisms and regulation. *Psychopharmacology* 2018; 235: 373–375.

98. Kryst J, Kawalec P, Mitoraj AM, *et al*. Efficacy of single and repeated administration of ketamine in unipolar and bipolar depression: a meta-analysis of randomized clinical trials. *Pharmacol Rep* 2020; 72: 543–562.

99. An D, Wei C, Wang J, *et al*. Intranasal ketamine for depression in adults: a systematic review and meta-analysis of randomized, double-blind, placebo-controlled trials. *Front Psychol* 2021; 12: 648691.

100. Marcantoni WS, Akoumba BS, Wassef M, *et al*. A systematic review and meta-analysis of the efficacy of intravenous ketamine infusion for treatment resistant depression: January 2009–January 2019. *J Affect Disord* 2020; 277: 831–841.

101. Majic T, Schmidt TT and Gallinat J. Peak experiences and the afterglow phenomenon: when and how do therapeutic effects of hallucinogens depend on psychedelic experiences? *J Psychopharmacol* 2015; 29: 241–253.

102. Sloshower J, Skosnik PD, Safi-Aghdam H, *et al*. Psilocybin-assisted therapy for major depressive disorder: an exploratory placebo-controlled, fixed-order trial. *J Psychopharmacol* 2023; 37: 698–706.

103. Nayak SM, Bradley MK, Kleykamp BA, *et al*. Control conditions in randomized trials of psychedelics: an ACTTION systematic review. *J Clin Psychiatry* 2023; 84: 22r14518.

104. Morris H. POD 52: Cinco sins of psychedelic science with Dr. Manoj Doss, https://www.patreon.com/posts/pod-52-cinco-of-69620551 (2022, accessed 18 September 2023).

105. Bradt J. Randomized controlled trials in music therapy: guidelines for design and implementation. *J Music Ther* 2012; 49: 120-149.

106. Garcia-Romeu A, Griffiths RR and Johnson MW. Psilocybin-occasioned mystical experiences in the treatment of tobacco addiction. *Curr Drug Abuse Rev* 2015; 7: 157–164.

107. Boutron I, Estellat C, Guittet L, *et al*. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 2006; 3: e425.

108. Nautiyal KM and Yaden DB. Does the trip matter? Investigating the role of the subjective effects of psychedelics in persisting therapeutic effects. *Neuropsychopharmacology* 2023; 48: 215–216.

109. Nicholas CR. Pilot RECAP study in healthy volunteers (RECAP), https://clinicaltrials.gov/ct2/show/NCT04842045 (2022, accessed 18 September 2023).

110. Lii TR, Smith AE, Flohr JR, *et al*. Trial of ketamine masked by surgical anesthesia in depressed patients. *medRxiv*, 2023.

111. Hrobjartsson A, Thomsen AS, Emanuelsson F, *et al*. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012; 344: e1119.

112. Even C, Siobud-Dorocant E and Dardennes RM. Critical approach to antidepressant trials. Blindness protection is necessary, feasible and measurable. *Br J Psychiatry* 2000; 177: 47–51.

113. Szigeti B, Nutt D, Carhart-Harris R, *et al*. On the fallibility of placebo control and how to address it: a case study in psychedelic microdosing. *PsyArXiv*, 2022.

114. Hartogsohn I. Set and setting, psychedelics and the placebo response: an extra-pharmacological perspective on psychopharmacology. *J Psychopharmacol* 2016; 30: 1259–1267.

115. Hartogsohn I. Constructing drug effects: a history of set and setting. *Drug Sci Policy Law* 2017; 3: 1–17.

116. Carhart-Harris RL, Kaelen M, Whalley MG, *et al*. LSD enhances suggestibility in healthy volunteers. *Psychopharmacology* 2015; 232: 785–794.

117. Gedin F, Blome S, Ponten M, *et al*. Placebo response and media attention in randomized clinical trials assessing cannabis-based therapies for pain: a systematic review and meta-analysis. *JAMA Netw Open* 2022; 5: e2243848.

118. Molendijk ML, Fried EI and van der Does W. The SMILES trial: do undisclosed recruitment practices explain the remarkably large effect? *BMC Med* 2018; 16: 243.

119. Ross N. 7. Political science. *New York Magazine*, 8 March 2022.

120. Wampold BE, Frost ND and Yulish NE. Placebo effects in psychotherapy: a flawed concept and a contorted history. *Psychol Conscious Theory Res Prac* 2016; 3: 108–120.

121. van Elk M and Yaden DB. Pharmacological, neural, and psychological mechanisms underlying psychedelics: a critical review. *Neurosci Biobehav Rev* 2022; 140: 104793.

122. Yaden DB and Griffiths RR. The subjective effects of psychedelics are necessary for their enduring therapeutic effects. *ACS Pharmacol Transl Sci* 2020; 4: 568–572.

123. Grieco SF, Castren E, Knudsen GM, *et al*. Psychedelics and neural plasticity: therapeutic implications. *J Neurosci* 2022; 42: 8439–8449.

124. Carhart-Harris RL and Friston KJ. REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacol Rev* 2019; 71: 316–344.

125. Feduccia A, Agin-Liebes G, Price CM, *et al*. The need for establishing best practices and gold standards in psychedelic medicine. *J Affect Disord* 2023; 332: 47–54.

126. Olson DE. Biochemical mechanisms underlying psychedelic-induced neuroplasticity. *Biochemistry* 2022; 61: 127–136.

127. Moncrieff J, Cooper RE, Stockmann T, *et al*. The serotonin theory of depression: a systematic umbrella review of the evidence. *Mol Psychiatry*. Epub ahead of print 20 July 2022. DOI: 10.1038/s41380-022-01661-0.

128. Baier AL, Kline AC and Feeny NC. Therapeutic alliance as a mediator of change: a systematic review and evaluation of research. *Clin Psychol Rev* 2020; 82: 101921.

129. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *J Am Stat Assoc* 1959; 54: 30–34.

130. Schiavone SR, Quinn KA and Vazire S. A consensus-based tool for evaluating threats to the validity of empirical research. *Psyarxiv*, 2023.

131. Notice of information on NIMH's considerations for research involving psychedelics and related compounds, https://grants.nih.gov/grants/guide/notice-files/NOT-MH-23-125.html (2022, accessed 18 September 2023).

132. Masters R and Houston J. *The varieties of psychedelic experience: the classic guide to the effects of LSD on the human psyche*. South Paris, ME: Park Street Press, 1966.